

\*\*\*\*\*  
WILEY  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 16 08:49:52 2000; MasPar time 24.34 Seconds  
Tabular output not generated. 672.401 Million cell updates/sec

Title: >US-09-360-125-1  
Description: (1-347) from US09360125.pep  
Perfect Score: 2538  
Sequence: 1 MAQSRDGGNPFAPSELDNP.....VRTAAANAAGAAENAFRAP 347

Scoring table: PAM 150  
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 49.533; Variance 128.029; scale 0.387

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length DB ID	Description	Pred. No.
1	2490	98.1	347 2 T08826	secretory carrier mem	0.00e+00
2	1299	51.2	338 2 S37395	secretory carrier mem	9.04e-181
3	748	29.5	324 2 T15283	hypothetical protein	1.57e-92
4	209	8.2	221 2 T09093	hypothetical protein	6.05e-12
5	141	5.6	1388 2 T00063	hypothetical protein	1.46e-03
6	129	5.1	1584 2 T00026	brain-specific angiot	3.15e-02
7	124	4.9	323 2 S20099	transforming protein	1.09e-01
8	125	4.9	436 2 E70387	conserved hypothetical	8.49e-02
9	116	4.6	302 2 H64871	oligopeptide transpor	7.42e-01
10	116	4.6	302 1 QREBOC	oligopeptide transpor	7.42e-01
11	118	4.6	357 2 A45619	ubiquinol--cytochrome	4.62e-01
12	116	4.6	437 2 S34959	NADH dehydrogenase (u	7.42e-01
13	116	4.6	440 2 T11319	NADH dehydrogenase su	7.42e-01
14	116	4.6	461 2 T10741	extensin-like protein	7.42e-01
15	117	4.6	822 2 I38728	epidermal growth fact	5.86e-01
16	118	4.6	892 2 A41697	nitrate assimilation	4.62e-01
17	115	4.5	287 2 A25877	cytochrome-c oxidase	9.39e-01
18	113	4.5	292 2 C64128	lic-1 protein B - Hae	1.50e-00
19	113	4.5	305 2 B33465	lic-1 protein B - Hae	1.50e-00
20	115	4.5	322 2 T13486	NADH dehydrogenase su	9.39e-01
21	113	4.5	431 2 H72247	preproteins translocas	1.50e-00
22	113	4.5	443 2 F30010	NADH dehydrogenase (u	1.50e-00
23	113	4.5	620 2 S06733	hydroxyproline-rich g	1.50e+00

24	115	4.5	760 2 T06291	extensin homolog T9B8	9.39e-01
25	114	4.5	948 2 A57640	retinoblastoma protei	1.19e+00
26	111	4.4	298 2 T13684	NADH dehydrogenase su	2.37e+00
27	112	4.4	373 2 S27266	secF protein - Strept	1.88e+00
28	112	4.4	475 2 A70320	nitrate transporter -	1.88e+00
29	111	4.4	496 2 T00691	hypothetical protein	2.37e+00
30	109	4.3	184 2 S78091	endocuticular protein	3.74e+00
31	108	4.3	232 2 A45612	H+-transporting ATP s	4.69e+00
32	110	4.3	265 1 E69254	cobalamin (5'-phospha	2.98e+00
33	110	4.3	299 2 A57652	cyclic nucleotide-gat	2.98e+00
34	109	4.3	421 2 S29599	acrosin (EC 3.4.21.1)	3.74e+00
35	110	4.3	439 2 C22845	NADH dehydrogenase (u	2.98e+00
36	110	4.3	448 2 T06076	extensin-like protein	2.98e+00
37	108	4.3	481 2 S47091	cyclase-associated pr	4.69e+00
38	110	4.3	530 1 G64918	phosphotransferase sy	2.98e+00
39	108	4.3	530 2 A45690	transactivator EBNA-2	4.69e+00
40	110	4.3	543 2 S25128	61K protein - Autogra	2.98e+00
41	110	4.3	543 2 H40781	hypothetical 60.7K pr	2.98e+00
42	109	4.3	705 2 S55420	conserved hypothetical	3.74e+00
43	109	4.3	804 2 S61395	probable Na+/H+-excha	3.74e+00
44	110	4.3	2715 2 T13049	evellid - fruit fly D	2.98e+00
45	109	4.3	3951 1 VFIHBI	F1 protein - avian in	3.74e+00

ALIGNMENTS

RESULT ENTRY TITLE DATE ORIGIN  
X  
T08826 #type complete  
secretory carrier membrane protein homolog propin1 - human  
#formal\_name Homo sapiens #common\_name man  
20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change  
20-Sep-1999  
T08826  
Z16482  
Winfield, S.L.; Tayebi, N.; Martin, B.M.; Ginns, E.I.;  
Sidransky, E.  
Genome Res. (1997) 7:1020-1026  
Identification of three additional genes contiguous to the  
glucocerebrosidase locus on chromosome 1q21: Implications  
for Gaucher Disease.  
#cross-references MUD:97474796  
#accession T08826  
#status translated from GB/EMBL/DBJ  
#molecule\_type DNA  
#residues 1-347 #label WIN  
#cross-references EMBL:AF023268; NID:g2564910; PID:g2564915

GENETICS  
#gene propin1  
#map\_position 1  
#introns  
SUMMARY  
length 347 #molecular-weight 38404 #checksum 7224  
Query Match 98.1%; Score 2490; DB 2; Length 347;  
Best local similarity 98.6%; Pred. No. 0.00e+00;  
Matches 342; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db	1	MAQSRDGGNPFAPSELDNP	QDPAVIOHRSROYATRDVYNP	ETREPPPPAYEPAPAP	60
Qy	1	MAQSRDGGNPFAPSELDNP	QDPAVIOHRSROYATLDVYNP	ETREPPPPAYEPAPAP	60
Db	61	LPPPSAPLSQPSRKLS	PTPEKNGYSYQASAAATAELLK	QBELNRKAEELDREREL	120
Qy	61	LPPPSAPLSQPSRKLS	PTPEKNGYSYQASAAATAELLK	QBELNRKAEELDREREL	120
Db	121	QHAALGTATQNNWPLP	SPFCVPQCFQDISMEIQFOE	ISMEIQFOEKTVMYILMWCST	180
Qy	121	QHAALGTATQNNWPLP	SPFCVPQCFQDISMEIQFOE	ISMEIQFOEKTVMYILMWCST	180
Db	181	NFLACLASFVETNNGAG	FGLSILWLLFTPCSFVCVYR	PMYKAFRSDSSNFNAFFNF	240
Qy	181	NFLACLASFVETNNGAG	FGLSILWLLFTPCSFVCVYR	PMYKAFRSDSSNFNAFFNF	240
Db	241	FDQDVLFLQAIQIPGNG	FGSGWISALVVPKGNATVSV	LMVLLVALLFTGTGI	300

```

241 FVQDVLFLQALIGPGWFSGISALVVPKGNATVSVLMLLVALFTGTAIVGLVMLKRI 300
Db 301 HSLYRTGASFOKAQOEFAAGVFNPAVTRTAANAAGAAENAFAP 347
QY 301 HSLYRTGASFOKAQOEFAAGVFNPAVTRTAANAAGAAENAFAP 347

RESULT 2
ENTRY #type complete
TITLE secretory carrier membrane protein 37 - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
09-Sep-1994
ACCESSIONS S37395; S37782
REFERENCE S37395
#authors Brand, S.H.; Castle, J.D.
#journal EMBO J. (1993) 12:3753-3761
#title SCAMP 37, a new marker within the general cell surface
recycling system.
#cross-references MUID:94008982
#accession S37395
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-338 #label BRA
#accession S37782
#molecule_type protein
#residues 53-78;105-116;299-322 #label BR2
KEYWORDS calcium binding; leucine zipper; transmembrane protein; zinc
finger
FEATURE
151-176 #domain transmembrane #status predicted #label TM1
222-239 #domain transmembrane #status predicted #label TM2
258-282 #domain transmembrane #status predicted #label TM3
SUMMARY #length 338 #molecular_weight 37998 #checksum 9593

Query Match 51.2%; Score 1299; DB 2; Length 338;
Best Local Similarity 55.3%; Pred. No. 9,04e-181;
Matches 187; Conservative 71; Mismatches 63; Indels 17; Gaps 14;

Db 6 SNPFADP-DLNNPFKDPSTQVTRNVP-PGLDEYNPF-S-DSR-T-PP-PGVKMPNVP 57
QY 8 GNPAEPSELONPFQDPAVQHRPSRQATLDVYAPFETREPPAYEPAPAPLPSPAP 67
Db 58 NTQPAI-WKPE--EHPAYTQITKEHALAQAEELKRQELERKAAELDREREMON--LS 112
QY 68 SLOPSRKLSPTEPKNGYSYQASAAA-ATAELLKKQELNKAELDRERELQHAALG 126
Db 113 QHG-RKNWNPPLPSNF-PVGCFCFYODFSVDIPVEFQKTVKMYLWMPHAYTLFLNIGC 170
QY 127 GTATRONWNPPLPS-FCPVQPCFFODISMEIQEFOKTVSTMYLWMCSTLALLNFLAC 185
Db 171 LAWFCDSSRAVDFGLSLWFLLELLPCSFVQVRYLYGAFRSDSFRFFVFFVYICQFA 230
QY 186 LASFCVEINNGAGFGLSLWLLFPFCFVQVRYLYGAFRSDSFRFFVFFVYICQFA 245
Db 231 VHVLOAAGFNWNGCWGTSSTGLNKNIPVGMIIIAALFTASAVISLWMPKHYGLYR 290
QY 246 LFVLOAIGIPGWFGWISALVVPKGNATVSVLMLLVALFTGTAIVGLVMLKRIHSLYR 305
Db 291 TTGASFEKAQOEFAAGVFNPAVTRTAANAAGAAENAFAP 347
QY 306 RTGASFEKAQOEFAAGVFNPAVTRTAANAAGAAENAFAP 347

RESULT 3
ENTRY #type fragment
TITLE hypothetical protein M01D7.2 - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
20-Sep-1999
ACCESSIONS T15283
```

```

REFERENCE 218322
#authors Gattung, S.; Goela, D.; Wilson, R.
#submission submitted to the EMBL Data Library, May 1997
#description The sequence of C. elegans cosmid M01D7.
#accession T15283
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues 1-324 #label GAT
#cross-references EMBL:AF003739; NID:g2105482; PID:g2105487;
#experimental_source strain Bristol N2; clone M01D7
GENETICS
#gene CESP:M01D7.2
#map_position 1
SUMMARY #length 324 #checksum 8640

Query Match 29.5%; Score 748; DB 2; Length 324;
Best Local Similarity 43.1%; Pred. No. 1.57e-92;
Matches 110; Conservative 70; Mismatches 63; Indels 12; Gaps 10;

Db 66 DLERRAQELRMREELDRRQSAAGNNLNTNAQNNAPRHPNPPPTIPIETPCYQDI 125
QY 98 ELLKQELNKAELDRRERELQHAALG-CT---ATRONWNPPLPSCFQVPCFQDI 152
Db 126 EVEIPVQFQKTVTFAYVFLMYLVLVNVLASL-FYMFAGSGIGQLFLACIQALFSP 184
QY 153 SMEIQEFOKTVSTMYLWMCSTLALLNFLACLSFCVETNNGAG-FGLSLWLLVFLTP 211
Db 195 CSFLFWFPVYKAFRNDSSFNFFVFLFFHCITFFV-OTLGLSNYA-CGHNALDTFN 242
QY 212 CSFVCWYRPMYKAFRSDSFRFFVFFVFFVQDVL-FVLQAIQPGWFGSGHVALVVK 270
Db 243 VSPVIALMLLSAIFET-VALTGMVTLVKVHRLYRGAGFSDKARQETNGVMSDAGVQ 301
QY 271 GNTAVSVLMLLVALFTGTAIVGLVYM-LKRIHSLYRRTGASFOKAQOEFAAGVFNPAVR 329
Db 302 RATQAATQAAGAAAF 316
QY 330 TAAANAAGAAENAF 344

RESULT 4
ENTRY #type complete
TITLE hypothetical protein F21B7.17 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
20-Sep-1999
ACCESSIONS T00903
REFERENCE 214208
#authors Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.;
Sun, H.; Conway, A.; Conway, A.; Kurtz, D.; Ojil, O.; Shen,
Y.K.; Toriumi, M.; Vysotskaia, V.; Yu, G.; Davis, R.W.;
Federapfel, N.A.; Theologis, A.; Ecker, J.R.
#submission submitted to the EMBL Data Library, January 1998
#description Genomic sequence for Arabidopsis thaliana BAC F21B7.
#accession T00903
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues 1-221 #label SHI
#cross-references EMBL:AC002560; NID:g2618677; PID:g2809248;
GSPDB:GN0059; ATSP:F21B7.17
GENETICS
#gene ATSP:F21B7.17
#map_position 1
#intons 18/3; 53/3; 74/3; 86/1; 117/2; 132/3; 155/2; 171/3; 193/3
SUMMARY #length 221 #molecular_weight 24765 #checksum 5742

Query Match 8.2%; Score 209; DB 2; Length 221;
Best Local Similarity 30.4%; Pred. No. 6.05e-12;
Matches 34; Conservative 38; Mismatches 36; Indels 4; Gaps 2;

Db 92 LSIIYFLAGVPGAVVLTWRPLRYRATRTDSALKFGAFFFFVFFHFAFCGFAVAPPVIFOG 151
```

\*\*\*\*\*

MSRCH\_PP

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_PP protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Jun 16 08:52:44 2000; MasPar time 23.36 Seconds  
Tabular output not generated. 664.234 Million cell updates/sec

Title: >US-09-360-125-3  
Description: (1-329) from US09360125.ppt  
Perfect Score: 2434  
Sequence: 1 MSADFTNPFADPVDVNPFDQ.....SSRTFHRAASAAQAFQGN 329

Scoring table: PAM 150  
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r62  
1: p1r1 2: p1r2 3: p1r3 4: p1r4

Statistics: Mean 49.328; Variance 126.753; scale 0.399

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1397	57.4	338	2 S37395	secretory carrier mem	5.93e-198
2	1223	50.2	347	2 T08826	secretory carrier mem	1.36e-169
3	670	27.5	324	2 T15283	hypothetical protein	7.18e-81
4	246	10.1	221	2 T00903	hypothetical protein	4.56e-17
5	128	5.3	684	2 T13491	NADH dehydrogenase	3.46e-02
6	129	5.3	731	2 T12226	NADH dehydrogenase	2.69e-02
7	126	5.2	700	2 T13471	NADH dehydrogenase	5.71e-02
8	124	5.1	686	2 T12125	NADH dehydrogenase	9.37e-02
9	124	5.1	698	2 T12560	NADH dehydrogenase	9.37e-02
10	122	5.0	285	2 S73885	MG135 homolog A65_Orf	1.53e-01
11	121	5.0	288	2 T13684	NADH dehydrogenase	1.96e-01
12	122	5.0	684	2 T12147	NADH dehydrogenase	1.53e-01
13	121	5.0	698	2 T12586	NADH dehydrogenase	1.96e-01
14	119	4.9	638	1 Q0B2M	mRNA maturase b14 - y	3.18e-01
15	119	4.9	736	2 T12114	NADH dehydrogenase	3.18e-01
16	119	4.9	740	2 T12223	NADH dehydrogenase	3.18e-01
17	118	4.8	383	2 C71863	rod shape-determining	4.04e-01
18	116	4.8	398	1 A22931	ubiquinol--cytochrome	6.51e-01
19	118	4.8	648	2 T12570	NADH dehydrogenase	4.04e-01
20	116	4.8	683	2 T12127	NADH dehydrogenase	6.51e-01
21	116	4.8	683	2 T12149	NADH dehydrogenase	6.51e-01
22	117	4.8	686	2 T12128	NADH dehydrogenase	6.51e-01
23	117	4.8	687	2 T12126	NADH dehydrogenase	5.13e-01

24	117	4.8	697	2 T13754	NADH dehydrogenase	5.13e-01
25	117	4.8	700	2 T13726	NADH dehydrogenase	5.13e-01
26	116	4.8	704	2 T13729	NADH dehydrogenase su	6.51e-01
27	116	4.8	734	2 T13685	NADH dehydrogenase	6.51e-01
28	117	4.8	739	2 T13770	NADH dehydrogenase	5.13e-01
29	114	4.7	196	2 S51908	cryptogene protein G1	1.04e-00
30	115	4.7	388	1 CB2M	ubiquinol--cytochrome	8.25e-01
31	115	4.7	495	1 S77406	protein kinase pknA (	8.25e-01
32	115	4.7	610	1 S30234	transcription factor	8.25e-01
33	115	4.7	685	2 T12129	NADH dehydrogenase	8.25e-01
34	114	4.7	704	2 T13581	NADH dehydrogenase su	1.04e-00
35	115	4.7	737	2 T12193	NADH dehydrogenase	8.25e-01
36	115	4.7	742	2 T13495	NADH dehydrogenase	8.25e-01
37	115	4.7	1407	2 B42239	adenylate cyclase (EC	8.25e-01
38	113	4.6	274	2 S42172	NADH dehydrogenase (u	1.32e-00
39	112	4.6	436	2 E70387	conserved hypochetica	1.67e-00
40	112	4.6	611	2 T13583	NADH dehydrogenase	1.67e-00
41	113	4.6	685	2 T12138	NADH dehydrogenase	1.32e-00
42	113	4.6	691	2 T12293	NADH dehydrogenase	1.32e-00
43	112	4.6	695	2 T12148	NADH dehydrogenase	1.67e-00
44	113	4.6	698	2 T12713	NADH dehydrogenase su	1.32e-00
45	113	4.6	732	2 T12194	NADH dehydrogenase	1.32e-00

ALIGNMENTS

RESULT 1

ENTRY  
TITLE  
ORGANISM  
DATE

S37395 #type complete  
secretory carrier membrane protein 37 - rat  
#formal\_name Rattus norvegicus #common\_name Norway rat  
31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change  
09-Sep-1994

ACCESSIONS

REFERENCE

#authors

#journal

#title

#cross-references

#accession

#status

#molecule\_type

#residues

#accession

#molecule\_type

#residues

KEYWORDS

FEATURE

151-176

222-239

258-282

SUMMARY

Query Match

Best Local Similarity

Matches

1 MSDFDNPNADPDLNPFDP

QY 1 MSADFTNPFADPVDVNP

DB 60 QPAINMPTTEHPAYTQIT

QY 59 QPVLQPSVEPTQPTQAV

DB 118 NNPPLPSNFPVPCFQD

QY 118 NNPPLPSNCPVKPCFQ

DB 178 SSRVDFGLSILWFL

QY 178 SSKGVDFGLSILWFL

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

SECRETARY CARRIER MEMBRANE PROTEIN 37 - RAT

```

Db 238 GFHNKNGCNWISSLTGLNKN-IPVGMIMIIITAAFTASAVISLVNFKVKGHYRTTGASF 296
Qy 238 GIPGLGDSGWIAALSTLDNLSAIVSVMMVAGFTLCAVLSVFLQRVHSLYRTTGASF 297
Db 297 EKAQGEFATGVMSNKTIVTAAANAASTA 324
Qy 298 QQAQEEFSQIGFSRTFTHRAASSAQA 325

RESULT 2
ENTRY #type complete
TITLE secretory carrier membrane protein homolog propin1 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
20-Sep-1999

ACCESSIONS
REFERENCE T08826
AUTHORS Winfield, S.L.; Tayebi, N.; Martin, B.M.; Ginns, E.I.;
Storansky, E.
JOURNAL Genome Res. (1997) 7:1020-1026
TITLE Identification of three additional genes contiguous to the
glucocorticoidase locus on chromosome 1q21: Implications
for Gaucher Disease.
CROSS-REFERENCES MUID:97474796
ACCESSION T08826
STATUS translated from GB/EMBL/DBJ
MOLECULE_type DNA
RESIDUES 1-347 #label WTN
CROSS-REFERENCES EMBL:AF032268; NID:g2564910; PID:g2564915
GENETICS
MAP_POSITION 1
INTRONS 22/3; 48/3; 130/1; 173/1; 226/2; 260/2; 299/3
SUMMARY #length 347 #molecular_weight 38404 #checksum 7224

Query Match 50.2%; Score 1223; DB 2; Length 347;
Best Local Similarity 52.6%; Pred. No. 1.36e-169;
Matches 175; Conservative 72; Mismatches 70; Indels 16; Gaps 12;

Db 9 NPFAPESELDNPFQDAVQIHRPQRYATRDVYNPFETREPPAPAPLPPPSAPS 68
Qy 7 NPFAADPVV-NPFQDPSVTQLTNAPQGLAE-FNPF-S-ETNAATTVPV-TQLPGSSQA 61
Db 69 -LQPSRLSPTEPNKYSYSQASAAATAELLKKQELNKAELDRRELQHALGG 127
Qy 62 VLQPS--VEPTQ-----T-P-QAVVSAQAQGLLQROEELDRKAAELERKRELQNT-VAN 112
Db 128 TATRONNPPLPSPQVPCPFODISMEIPOEFQKTVSTMYLWMCSTLALLNPLACLA 187
Qy 113 LHVRONNPPLPSPWCPKPCYQDFSTEIPADYQRIKRMVLYLWMLHSVTILFLLNLA 172
Db 188 SFCVETNNGAGFGLSILWLLFTPCSFVCTWRPMYKAFRSDSSNFNFAFFNFDDQVLF 247
Qy 173 WFSNNSKGVDFGLSILWLLFTPCSFVCTWRPMYKAFRSDSSNFNFAFFNFDDQVLF 232
Db 248 VLQAGIPGWCFSQWISALVVPKNT-AVSVMLLVALLFTGIVLGMKRLKHSLYR 306
Qy 233 IQLVGLPGLGDSWIAALSTLDNLSAIVSVMMVAGFTLCAVLSVFLQRVHSLYR 292
Db 307 TGASFQQAQEEFAAGVFSNPVTRTAANAAGA 339
Qy 293 TGASFQQAQEEFSQIGFSRTFTHRAASSAQA 325

RESULT 3
ENTRY #type fragment
TITLE T15283
ORGANISM hypothetical protein M01D7.2 - Caenorhabditis elegans
(formragment)
#formal_name Caenorhabditis elegans
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
20-Sep-1999
ACCESSIONS T15283

```

```

REFERENCE
AUTHORS Gattung, S.; Goela, D.; Wilson, R.
SUBMISSION submitted to the EMBL Data Library, May 1997
DESCRIPTION The sequence of C. elegans cosmid M01D7.
ACCESSION T15283
STATUS Preliminary; translated from GB/EMBL/DBJ
MOLECULE_type DNA
RESIDUES 1-324 #label GAT
CROSS-REFERENCES EMBL:AF003739; NID:g2105482; PID:g2105487;
PIDN:AAB58069.1; GSPDB:GN00019; GESP:M01D7.2
EXPERIMENTAL_source strain Bristol N2; clone M01D7
GENETICS
GENE CESP:M01D7.2
MAP_POSITION 1
SUMMARY #length 324 #checksum 8640

Query Match 27.5%; Score 670; DB 2; Length 324;
Best Local Similarity 36.9%; Pred. No. 7.18e-81;
Matches 109; Conservative 71; Mismatches 102; Indels 13; Gaps 12;

Db 25 DFNPFANRAGSNQPTTHQSTGLNKSAGMDLFRKQDQLERRAQE-LRMREEELDR 83
Qy 36 EFNPFSE-TNAATTVPV-TQLPGSSQPAVLQPSVEPTQTPQAVVSAQAQGLLQROEELDR 94
Db 84 RORSAA-GGNLNTNAQNAPRPHNPPLPIIPIEPCFYQDIEVEIPVQKQTFYFY 142
Qy 95 KAELEKERELEQNTVANLHVQRNNPPPLSPWCPKCYQDFSTEIPADYQRIKMLY 154
Db 143 VFLMYVLAVVNLASLIFYMIFAG-GSIGOLF-LACIOLALFSPCSFELFPRPVYKAFRN 200
Qy 155 LWMLHSVTILFLLNLAQAW--FSGNSKGVDFGLSILWLLFTPCFCLWTRPIYKAFRS 212
Db 201 DSSNFVFFVFLFHC-ITFTV-QTLGLSNYA-CGWINALDTF-NVSVPIALLMLISAI 256
Qy 213 DNSFFVFFVFFV-FF-CQIGIYIQLVGPGLGDSGWIAALSTLDNLSAIVSVMMVAG 270
Db 257 TETVALGTMVTLVYVHLYRGAGFSDIKARQETNGVNSDAGVQRTAQATQAA 311
Qy 271 FFTLCVLSVFLQRVHSLYRTRTGASFQQAQEEFSQIGFSRTFTHRAASSAQA 325

RESULT 4
ENTRY #type complete
TITLE hypothetical protein F21B7.17 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
20-Sep-1999
ACCESSIONS T00903
REFERENCE Z14208
AUTHORS Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.;
Sun, H.; Conway, A.; Conway, A.; Kurtz, D.; Oji, O.; Shen,
Y.K.; Toriumi, M.; Vysotskaia, V.; Yu, G.; Davis, R.W.;
Fiederspiel, N.A.; Theologis, A.; Ecker, J.R.
SUBMISSION submitted to the EMBL Data Library, January 1998
DESCRIPTION Genomic sequence for Arabidopsis thaliana BAC F21B7.
ACCESSION T00903
STATUS Preliminary; translated from GB/EMBL/DBJ
MOLECULE_type DNA
RESIDUES 1-221 #label SHI
CROSS-REFERENCES EMBL:AC002560; NID:g2618677; PID:g2809248;
GSPDB:GN0059; ATSP:F21B7.17
GENETICS
GENE ATSP:F21B7.17
MAP_POSITION 1
INTRONS 18/3; 53/3; 74/3; 86/1; 117/2; 132/3; 155/2; 171/3; 193/3
SUMMARY #length 221 #molecular_weight 24765 #checksum 5742

Query Match 10.1%; Score 246; DB 2; Length 221;
Best Local Similarity 33.3%; Pred. No. 4.56e-17;
Matches 38; Conservative 37; Mismatches 32; Indels 7; Gaps 4;

Db 92 LSIIVFLAGVPGAVLVLRPLRYRATRTDSALKFGAFFFFFYVFIHAFCGFAA-VAPPVIFQ 150

```

\*\*\*\*\*  
M P S R L H  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd.  
MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Jun 16 08:48:16 2000; MasPar time 15.54 Seconds  
Tabular output not generated. 680.011 Million cell updates/sec

Title: >US-09-360-125-1  
Description: (1-347) from US09360125.pep  
Perfect Score: 2538  
Sequence: 1 MAQSDGNGNPAEPSELDNP.....VNTAAANAAGAAENAFRAP 347

Scoring table: PAM 150  
Gap 11  
Searched: 83857 seqs, 30454973 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot38  
1:swissprot

Statistics: Mean 50.311; Variance 114.866; scale 0.438  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES															
Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.								
1	2490	98.1	347	1	SCA3_HUMAN	SECRETORY CARRIER-ASSO	0.00e+00								
2	2321	91.4	349	1	SCA3_MOUSE	SECRETORY CARRIER-ASSO	0.00e+00								
3	1299	51.2	338	1	SCAL_RAT	SECRETORY CARRIER-ASSO	1.11e-20								
4	1279	50.4	338	1	SCAL_HUMAN	SECRETORY CARRIER-ASSO	5.11e-20								
5	1256	49.5	328	1	SCA2_HUMAN	SECRETORY CARRIER-ASSO	8.29e-19								
6	124	4.9	323	1	JUND_CHICK	TRANSCRIPTION FACTOR J	1.91e-02								
7	125	4.9	410	1	HID_DROME	HEAD INVOLUTION DEFECT	1.45e-02								
8	119	4.7	579	1	YN05_CAEEL	PUTATIVE SODIUM- AND C	7.35e-02								
9	116	4.6	302	1	OPPC_ECOLI	OLIGOPEPTIDE TRANSPORT	1.62e-01								
10	116	4.6	302	1	OPPC_SALTY	OLIGOPEPTIDE TRANSPORT	1.62e-01								
11	117	4.6	822	1	EPS8_HUMAN	EPIDERMAL GROWTH FACTO	1.25e-01								
12	118	4.6	892	1	NIRA_EMENT	NITROGEN ASSIMILATION	9.59e-02								
13	113	4.5	292	1	LICB_HAEIN	LICB PROTEIN.	3.54e-01								
14	113	4.5	620	1	EXTN_TOBAC	EXTENSIN PRECURSOR (CE	3.54e-01								
15	114	4.5	3110	1	HD_RAT	HUNTINGTIN (HUNTINGTON	2.74e-01								
16	112	4.4	373	1	SECF_STRCO	PROTEIN-EXPORT MEMBRAN	4.58e-01								
17	111	4.4	496	1	YML5_ARATH	HYPOTHETICAL MLO-LIKE	5.91e-01								
18	111	4.4	527	1	RBFL_CANAL	TRANSCRIPTION FACTOR R	5.91e-01								
19	112	4.4	899	1	PMAL_KLOLA	PLASMA MEMBRANE ATPASE	4.58e-01								
20	108	4.3	175	1	Y433_METJA	HYPOTHETICAL PROTEIN M	1.26e+00								
21	108	4.3	305	1	NULM_APIILI	NADH-UBIQUINONE OXIDOR	1.26e+00								
22	110	4.3	464	1	GBA_LEIDO	GUANINE NUCLEOTIDE-BIN	7.62e-01								
23	108	4.3	481	1	CAP_CHLYR	ADENYLYL CYCLASE-ASSOC	1.26e+00								

RESULT 1  
ID SCA3\_HUMAN STANDARD; PRT; 347 AA.  
AC O14828; 015128;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DE 15-DEC-1998 (Rel. 37, Last annotation update)  
DE SECRETORY CARRIER-ASSOCIATED MEMBRANE PROTEIN 3.  
GN SCAMP3 OR PROPIN1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP TISSUE=BRNIN;  
RC SEQUENCE FROM N.A.  
RX MEDLINE; 97474796.  
RA Winfield S.L., Tayebi N., Martin B.M., Ginns E.I., Sidransky E.;  
RT Identification of three additional genes contiguous to the  
RT glucocerebrosidase locus on chromosome 1q21: implications for Gaucher  
RT disease.";  
RL Genome Res. 7:1020-1026(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98017831.  
RA Singleton D.R., Wu T.T., Castle J.D.;  
RT Three mammalian SCAMPs (secretory carrier membrane proteins) are  
RT highly related products of distinct genes having similar subcellular  
RT distributions.";  
RL J. Cell Sci. 110:2099-2107(1997).  
CC -1- FUNCTION: FUNCTIONS IN POST-GOLGI RECYCLING PATHWAYS. ACTS AS A  
CC RECYCLING CARRIER TO THE CELL SURFACE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN  
CC HEART AND SKELETAL MUSCLE.  
CC -1- SIMILARITY: BELONGS TO THE SCAMP FAMILY.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/  
CC or send an email to licensed@isb-sib.ch).  
CC -----  
CC EMBL; AF023268; AAC51821.1; -  
CC DR EMBL; AF005039; AAB62724.1; -  
CC KW Transmembrane; Transport; Protein transport; Multigene family.  
CC TRANSMEM 171 191 POTENTIAL.

ALIGNMENTS

24 108 4.3 485 1 SSGP\_VOLCA SULFATED SURFACE GLYC 1.26e+00  
25 109 4.3 491 1 Y039\_HUMAN HYPOTHETICAL PROTEIN K 9.81e-01  
26 110 4.3 497 1 NUD4M\_ECACA NADH-UBIQUINONE OXIDOR 1.26e+00  
27 110 4.3 530 1 PTOA\_ECOLI PTS SYSTEM, MALTOSE AN 7.62e-01  
28 110 4.3 543 1 VP61\_NPVAC 61 KD PROTEIN. 7.62e-01  
29 109 4.3 705 1 YWJF\_BACSU HYPOTHETICAL 79.2 KD P 9.81e-01  
30 108 4.3 815 1 MK07\_HUMAN MITOGEN-ACTIVATED PROT 1.26e+00  
31 108 4.3 947 1 PMA2\_YEAST PLASMA MEMBRANE ATPASE 1.26e+00  
32 108 4.3 3951 1 VGF1\_IBVB F1 PROTEIN. 9.81e-01  
33 106 4.2 125 1 SECG\_BORBU PROBABLE PROTEIN-EXPO 2.07e+00  
34 106 4.2 257 1 YXP2\_XANCP HYPOTHETICAL 26.9 KD P 2.07e+00  
35 107 4.2 389 1 NDPP\_MOUSE NPC DERIVED PROLINE RI 1.62e+00  
36 106 4.2 389 1 FL\_ORISA PUTATIVE TRANSCRIPTION 2.07e+00  
37 107 4.2 412 1 ALF\_PETHY ALF PROTEIN (ABERRANT 1.62e+00  
38 107 4.2 426 1 EXLP\_TOBAC PISTIL-SPECIFIC EXTENS 1.62e+00  
39 107 4.2 443 1 HXA3\_HUMAN HOMEBOX PROTEIN HOX-A 1.62e+00  
40 106 4.2 444 1 NUD4M\_LOCM1 NADH-UBIQUINONE OXIDOR 2.07e+00  
41 106 4.2 473 1 NORB\_PSEST NITRIC-OXIDE REDUCTASE 2.07e+00  
42 107 4.2 551 1 YD25\_YEAST HYPOTHETICAL 62.6 KD P 1.62e+00  
43 106 4.2 635 1 VP40\_HSV11 CAPSID PROTEIN P40 (VI 1.62e+00  
44 107 4.2 1744 1 TENS\_CHICK TENSIN. 1.62e+00  
45 105 4.1 364 1 FTSW\_BORBU CELL DIVISION PROTEIN 2.65e+00



301 QEEFSQIFSSRTFHRASSAAQGAFOGN 329

```
60 QPAIMKPTTEHPAYQTITKEHALAQAEELLKRQEELEKKAELDRREREMQN-LSQ-HGRK 117
```

FT	CONFLICT	119	119	N -> I (IN REF. 2).
FT	CONFLICT	136	136	D -> E (IN REF. 2).
FT	CONFLICT	247	247	I -> W (IN REF. 2).

FT TRANSMEM 197 217 POTENTIAL.  
FT TRANSMEM 247 267 POTENTIAL.  
FT TRANSMEM 277 297 POTENTIAL.  
FT CONFLICT 3 3 O -> R (IN REF. 2).  
FT CONFLICT 38 38 R -> L (IN REF. 2).  
FT CONFLICT 74 74 K -> M (IN REF. 2).  
FT CONFLICT 235 235 A -> V (IN REF. 2).  
FT CONFLICT 239 239 N -> I (IN REF. 2).  
FT CONFLICT 242 242 D -> V (IN REF. 2).  
FT CONFLICT 331 331 R -> A (IN REF. 2).  
SQ SEQUENCE 347 AA; EC55F3793D20E58C CRC64;

Query Match 98.18; Score 2490; DB 1; Length 347;  
Best Local Similarity 98.6%; Pred. No. 0.00e+00;  
Matches 342; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1 MAQRDGGNPFAPSELDNPFQDPAVQHRSQYATLDVYNPFPETREPPAYEPAPAP 60  
QY 1 MAQRDGGNPFAPSELDNPFQDPAVQHRSQYATLDVYNPFPETREPPAYEPAPAP 60

Db 61 LPPPSAPSLQSRKLSPTPEKNGYSYQASAAAATAELLKKQELNKAELDRREL 120  
QY 61 LPPPSAPSLQSRKLSPTPEKNGYSYQASAAAATAELLKKQELNKAELDRREL 120

Db 121 QHAALGGTATRONNWPPLPSPFCVQPCFQDIMEIPQEFQKTVSTMYLWMCSTLALL 180  
QY 121 QHAALGGTATRONNWPPLPSPFCVQPCFQDIMEIPQEFQKTVSTMYLWMCSTLALL 180

Db 181 NFLACLASFCVETNNGAGFGLSILWLLFTPCSFVQWYRPMYKAFRSDSSFNFAFFNF 240  
QY 181 NFLACLASFCVETNNGAGFGLSILWLLFTPCSFVQWYRPMYKAFRSDSSFNFAFFNF 240

Db 241 FDQDVLFLVQAIQIPGFGSGWISALVYKNGTAVSVLMLLVALFTGIAVLGIVMLKRI 300  
QY 241 FDQDVLFLVQAIQIPGFGSGWISALVYKNGTAVSVLMLLVALFTGIAVLGIVMLKRI 300

Db 301 HSLYRTGASFOKAQOQFAAGVSNPAVTRTAANAAGAAENAFRAP 347  
QY 301 HSLYRTGASFOKAQOQFAAGVSNPAVTRTAANAAGAAENAFRAP 347

RESULT 2  
ID SCAL\_RAT STANDARD; PRT; 349 AA.  
AC Q35609;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DE SECRETORY CARRIER-ASSOCIATED MEMBRANE PROTEIN 3.  
GN SCAMP3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
[1]  
SEQUENCE FROM N.A.  
MEDLINE: 98017831.  
RA Singleton D.R., Wu T.T., Castle J.D.;  
RT Three mammalian SCAMPs (secretory carrier membrane proteins) are highly related products of distinct genes having similar subcellular distributions.  
RL J. Cell Sci. 110:2099-2107(1997).  
CC -1- FUNCTION: FUNCTIONS IN POST-GOLGI RECYCLING PATHWAYS. ACTS AS A RECYCLING CARRIER TO THE CELL SURFACE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE SCAMP FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: L202079; ; NOT ANNOTATED.CDS  
DR Transmembrane; Transport; Protein transport; Multigene family.  
FT TRANSMEM 156 176 POTENTIAL.  
FT TRANSMEM 182 202

DR EMBL; AF005036; AAB62721.1; --  
KW Transmembrane; Transport; Protein transport; Multigene family.  
FT TRANSMEM 169 189 POTENTIAL.  
FT TRANSMEM 200 220 POTENTIAL.  
FT TRANSMEM 236 256 POTENTIAL.  
FT TRANSMEM 277 297 POTENTIAL.  
SQ SEQUENCE 349 AA; 38387 MW; D54A514769A49309 CRC64;

Query Match 91.48; Score 2321; DB 1; Length 349;  
Best Local Similarity 89.4%; Pred. No. 0.00e+00;  
Matches 312; Conservative 23; Mismatches 12; Indels 2; Gaps 1;

Db 1 MAQRDGTGNPSPDSGLDNPFQDPAVQHRSQYATLDVYNPFPENREPPAYEPAPAP 60  
QY 1 MAQRDGTGNPSPDSGLDNPFQDPAVQHRSQYATLDVYNPFPENREPPAYEPAPAP 60

Db 61 APPLPPSAPSVQSRKLSPTPEPNYGSYSTQASAAAATAELLKKQELNKAELDRRER 120  
QY 61 --LPPPSAPSVQSRKLSPTPEPNYGSYSTQASAAAATAELLKKQELNKAELDRRER 118

Db 121 ELQHVAGGAGTATRONNWPPLPSPFCVQPCFQDIMEIPQEFQKTVSTMYLWMCSTLAL 180  
QY 121 ELQHVAGGAGTATRONNWPPLPSPFCVQPCFQDIMEIPQEFQKTVSTMYLWMCSTLAL 178

Db 181 LNFACLASFCVETNNGAGFGLSILWLLFTPCSFVQWYRPMYKAFRSDSSFNFAFFNF 240  
QY 179 LNFACLASFCVETNNGAGFGLSILWLLFTPCSFVQWYRPMYKAFRSDSSFNFAFFNF 238

Db 241 IFVQDVLFLVQAIQIPGFGSGWISALVYKNGTAVSVLMLLVALFTGIAVLGIVMLK 300  
QY 239 IFVQDVLFLVQAIQIPGFGSGWISALVYKNGTAVSVLMLLVALFTGIAVLGIVMLK 298

Db 301 RIHSLYRTGASFOKAQOQFAAGVSNPAVTRTAANAAGAAENAFRAP 349  
QY 299 RIHSLYRTGASFOKAQOQFAAGVSNPAVTRTAANAAGAAENAFRAP 347

RESULT 3  
ID SCAL\_RAT STANDARD; PRT; 338 AA.  
AC P56603;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE SECRETORY CARRIER-ASSOCIATED MEMBRANE PROTEIN 1 (SCAMP 37).  
GN SCAMP1 OR SCAMP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
[1]  
SEQUENCE FROM N.A.; AND PARTIAL SEQUENCE.  
RP STRAIN-SPRAGUE-DAWLEY; TISSUE=BRAIN;  
RX MEDLINE: 94008982.  
RA Brand S.H., Castle J.D.;  
RT "SCAMP 37, a new marker within the general cell surface recycling system."  
RL EMBO J. 12:3753-3761(1993).  
CC -1- FUNCTION: FUNCTIONS IN POST-GOLGI RECYCLING PATHWAYS. ACTS AS A RECYCLING CARRIER TO THE CELL SURFACE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE SCAMP FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: L202079; ; NOT ANNOTATED.CDS  
DR Transmembrane; Transport; Protein transport; Multigene family.  
FT TRANSMEM 156 176 POTENTIAL.  
FT TRANSMEM 182 202